FACT SHEET: iRODS integrated Rule Oriented Data System

Open Source Data Grid, Helping People Organize and Manage
Large Collections of Distributed Digital Data

Key Features

Build sharable collections in iRODS from data distributed across file systems and tape archives.

iRODS iCAT Metadata Catalog stores state information and descriptive metadata in a database, enabling search, management, controlling and tracking of all data access and manipulation.

Rule Engine applies user-defined Policies and Rules to data to automate administrative tasks, enforce management policies, and evaluate assessment criteria, enabling large-scale collections.

Versatile and Configurable

Data Grid or "Intelligent Cloud" for sharing data. Logical namespace spans administrative domains, institutions, and international collaborations. Widely used in research community: *Ocean Observatories Initiative (OOI); Temporal Dynamics of Learning Center (TDLC); Large Synoptic Survey Telescope (LSST); Southern California Earthquake Center (SCEC); High Performance Computing (NSF TeraGrid, NASA Center for Computational Sciences), many others.* **Independent Evaluations** at *irods.org/index.php/Publications*.

Preservation Environments support full life cycle management and long-term accessibility of electronic records in **Trustworthy Repositories**: *National Archives Transcontinental Persistent Archives Prototype (NARA TPAP), Distributed Custodial Archival Preservation Environments.*

Digital Library configurations support large-scale publication: *French National Library.*

Scalable collections: From personal laptop to institutional repositories to international projects with thousands of users and petabytes of data in hundreds of millions of files at sites around the world. >4 PB managed worldwide. Parallel data transfers at ~70% of available network bandwidth.

Federation of multiple iRODS Data Grids enables **flexible architectures:** central archives, master-slave data grids, chained data grids, and deep archives, through independent but collaborating iCAT Metadata Catalogs.
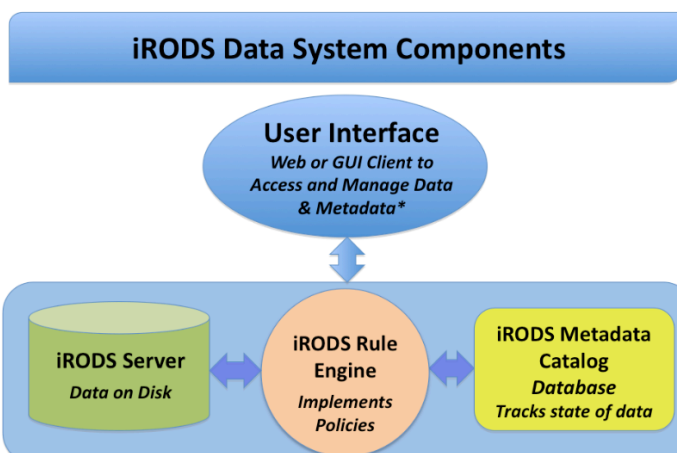
iRODS interfaces (GUI, Web, WebDAV, command line) let users search for, access and view, add/extract metadata, annotate, analyze and process, manage, replicate, copy, share, repurpose, control and track access, subscribe, and more.

iRODS Server software and **Rule Engine** run on each data server. The **iRODS iCAT Metadata Catalog** uses a database to track metadata describing data and everything that happens to it

User and Community-defined Policies are mapped to computer-actionable Rules applied at each storage system.

Procedures are mapped to workflows composed by chaining Micro-services (C code encapsulating a desired function). Workflows are executed at storage locations, under Rule control.

State information in iCAT Catalog tracks outcome of each Micro-service, can be queried to validate assessment criteria such as data authenticity. Access controls and audit trails can be analyzed to verify policy enforcement, enabling Trustworthy Repositories.

Highly extensible through user-defined Rules and Micro-services. Policies can be added without modifying the core code. API Interfaces: C, Unix shell, Java, Python, Kepler, Taverna, Web.

Interoperable with other data management systems e.g. Fedora, Dspace.

Open source BSD license. Quick and easy install.

System Requirements: *Hardware agnostic:* can support most storage (filesystems, tape, database). *iRODS Server:* Linux, Solaris, Macintosh, AIX, Windows. *iCAT Metadata Catalog:* open source PostgreSQL (optional install in iRODS install package), Oracle, MySQL.

Community: The nonprofit Data Intensive Cyberinfrastructure Foundation is the home of the growing iRODS open source community, which contributes new code for the benefit of all.

Development partners: CC-IN2P3; EU SHAMAN; Australian ARCS; UK e-Science; King's College; and others.

Recognition: *Internet2 IDEA Award* for innovation enabling "transformational progress in research, teaching and learning." *Society of American Archivists Jameson Award* for addressing "major challenges of managing, preserving, and providing access to electronic records."

DICE Center: Growing out of 10+ years of user-driven development, iRODS is a project of the Data Intensive Cyber Environments Center at UNC, with core development at INC UCSD. The DICE Center is affiliated with UNC's School of Information and Library Science and RENCI.

Funding: iRODS core development is funded by the National Archives and Records Administration and the National Science Foundation.
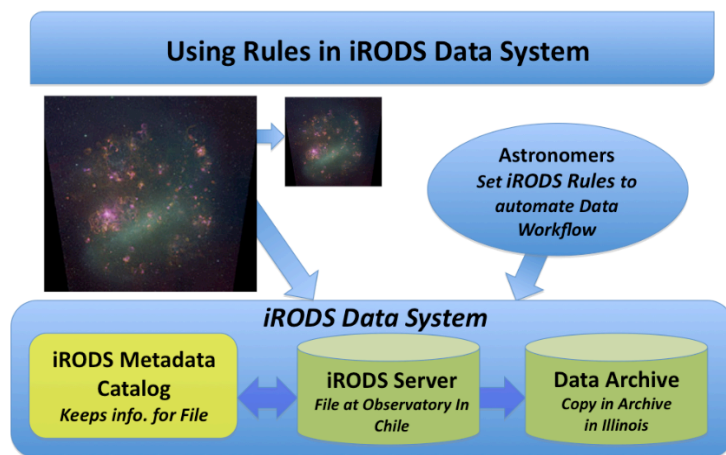
Rules for Automated Workflows

When astronomers take photographs in a sky survey and add to a data collection, they can set iRODS Rules to automatically extract descriptive metadata and record it in iRODS Metadata Catalog, create a thumbnail image for a Web-based gallery, replicate copies to other repositories for backup, and run an analysis program to identify related images. This **distributed workflow automation** makes management feasible for today's terabyte and petabyte scale data collections.